AD-A088 072 PRINCETON UNIV NJ DEPT OF STATISTICS F/6 12/1
FORMULAS FOR A TWO-SAMPLE MONTE CARLO SWINDLE.(U)
NOV 79 K KAFADAR DAAG29-76-G-0298
UNCLASSIFIED TR-153-SER-2 ARO-14244.13-M NL

1 OF 1
AD
A088072

END
DATE
FILMED
9-80
DTIC

18 ARO

AD A088072

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: 19 14244.13-M	2. GOVT ACCESSION NO.: AD-A088 072	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): 6 FORMULAS FOR A TWO-SAMPLE MONTE CARLO SWINDLE.		5. TYPE OF REPORT & PERIOD COVERED: Technical rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): 10 Karen Kafadar		8. CONTRACT OR GRANT NUMBER(s): DAAG29-76-G-0298
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Princeton University, Princeton, NJ 08540		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: U. S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709		12. REPORT DATE: 11 Nov 79
		13. NUMBER OF PAGES: 20
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 13		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

14 77-153-EER-2 15

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC ELECTE AUG 18 1980 A

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

probability
Monte Carlo methods
mathematical formulas
simulation
methodology
sampling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In order to obtain accurate estimates of the percentage points of the two-sample "t"-like statistic, we use a Monte Carlo swindle based on the method of conditional probabilities. The formulas for this implementation are described in this report.

DD FORM 1473, 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406873

80 8 14 113

DDC FILE COPY

FORMULAS FOR A TWO-SAMPLE MONTE CARLO SWINDLE

by

Karen Kafadar
Princeton University

Technical Report No. 153, Series 2
Department of Statistics
Princeton University
November 1979

This research was supported in part by a contract with the U.S. Army Research Office, No. DAAG29-76-0298, awarded to the Department of Statistics, Princeton University, Princeton, New Jersey.

A B S T R A C T

In order to obtain accurate estimates of the percentage points of the two-sample "t"-like statistic, we use a Monte Carlo swindle based on the method of conditional probabilities. The formulas for this implementation are described in this report.

0. INTRODUCTION

Efficient methods of problem solving using the computer are necessary for both accuracy and speed. These requirements often depend upon partial reduction of the problem to an analytic solution. This is the goal of Monte Carlo simulation, as opposed to experimental sampling.

This report is an extension of the formulas derived in Gross ([1]) for estimating the tail distribution of a statistic. In this work, the statistic was the form of a Student's-t for one sample. Here we derive the corresponding formulas for the two-sample statistic.

A. Location swindle.

Let $X_1,\ldots,X_{n_1}$ be a (pseudo-) random sample from population 1 having location parameter μ_1, and let $x_1,\ldots,x_{n_2}$ be a (pseudo-) random sample from population 2 having location parameter μ_2. Following the generation method described in [8], each X_i is generated as Z_i/Y_i, where $Z_i \sim N(0,1)$, and Y_i comes from a positive distribution. Let $A = \sum_{i=1}^{n_1} X_i Y_i^2 / \sum_{i=1}^{n_1} Y_i^2$. Likewise, each x_i is generated as z_i/y_i, where $z_i \sim N(0,1)$, y_i comes from a positive distribution, and $a = \sum_{i=1}^{n_2} x_i y_i^2 / \sum_{i=1}^{n_2} y_i^2$. Notice that, conditional on $\{Y_i\}$ (respectively, $\{y_i\}$), A (resp. a) $\sim N(0, 1/\sum Y_i^2)$ (resp. $N(0, 1/\sum y_i^2)$).

Let $T(\underline{X})$ (resp. $t(\underline{x})$) be an estimate of μ_1 (respec-

tively, μ_2), and let S_T (resp. s_t) be an estimate of the width of the distribution of $T(\underline{X})$ (resp. $t(\underline{x})$). Finally, let $\underline{C} = (\underline{X} - A)$, and $\underline{c} = (\underline{x} - a)$. Note that since width estimates are generally location invariant (as ours will be here), $S_T = S_{T(\underline{C})}$, $s_t = s_{t(\underline{c})}$, and since location estimates are translation invariant, $T(\underline{X}) = T(\underline{C}) + A$, $t(\underline{x}) = t(\underline{c}) + a$. In addition, we will assume that T and t are *symmetric* functions.

In order to use

$$W = \frac{T(\underline{X}) - t(\underline{x})}{\sqrt{S_T^2 + s_t^2}}$$

to test the null hypothesis H_0: $\mu_1 = \mu_2$ and subsequently to derive a confidence interval for $\mu_1 - \mu_2$, we determine the percentage points from the distribution of W as follows:

$$\begin{aligned}
P_0\{W \le k\} &= E_0[I_{(-\infty,k]}(W)] \\
&= E_0[I_{(-\infty,k]}((T(\underline{X}) - t(\underline{x}))/(S_T^2 + s_t^2)^{1/2})] \\
&= E_0\left[I_{(-\infty,k]}\left[\frac{(T(\underline{C})+A) - (t(\underline{c})+a)}{(S_{T(C)}^2 + s_{t(c)}^2)^{1/2}}\right]\right] \\
&= E_{\underline{C},\underline{c},\underline{Y},\underline{y}}\, E_0\left[I_{(-\infty,k]}\left[\frac{(A-a) + (T(\underline{C}) - t(\underline{c}))}{(S_{T(C)}^2 + s_{t(c)}^2)^{1/2}}\right] \,\middle|\, \underline{C},\underline{c},\underline{Y},\underline{y}\right] \\
&= E_{\underline{C},\underline{c},\underline{Y},\underline{y}}\, P_0\left[\frac{(A-a) + (T(\underline{C}) - t(\underline{c}))}{(S_{T(C)}^2 + s_{t(c)}^2)^{1/2}} \le k \,\middle|\, \underline{C},\underline{c},\underline{Y},\underline{y}\right] \\
&= E_{\underline{C},\underline{c},\underline{Y},\underline{y}}\, P_0[(A-a) \le k(S_{T(C)}^2 + s_{t(c)}^2)^{1/2} - T(\underline{C}) + t(\underline{c}) \mid \underline{C},\underline{c},\underline{Y},\underline{y}]
\end{aligned}$$

$$= \underset{\underline{C},\underline{c},\underline{X},\underline{x}}{E} \Phi \left[\frac{k(S^2_{T(\underline{C})} + s^2_{t(\underline{c})})^{1/2} - T(\underline{C}) + t(\underline{c})}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} y_i^2)^{1/2}} \right]$$

$$\stackrel{\Delta}{=} \frac{1}{N} \sum_{j=1}^{N} \Phi \left[\frac{k(S_j^2 + s_j^2)^{1/2} - (T_j - A_j) + (t_j - a_j)}{(1/\sum_{i=1}^{n_1} Y_{ij}^2 + 1/\sum_{i=1}^{n_2} y_{ij}^2)^{1/2}} \right] .$$

where N is the number of samples from each of populations 1 and 2. (The subscripts on "E" refer to the variables over which the expectation is conditioned.)

Denote the term being summed as Y_j, which in this case is

$$\Phi_{1j} = \Phi \left[\frac{k(S_j^2 + s_j^2)^{1/2} - (T_j - A_j) + (t_j - a_j)}{(1/\sum_{i=1}^{n_1} Y_{ij}^2 + 1/\sum_{i=1}^{n_2} y_{ij}^2)^{1/2}} \right] .$$

Then $P_0\{W \le k\}$ can be estimated unbiasedly by

$$P_0 = \frac{1}{N} \cdot \sum_{j=1}^{n} Y_j ,$$

the variance of which ($= E(P_0^2) - [E(P_0)]^2$) can be estimated by

$$\mathrm{Var}(P_0) \stackrel{\Delta}{=} \frac{1}{N}\left[\frac{1}{N} \sum_{j=1}^{N} Y_j^2 - (P_0)^2 \right]$$

The next step to consider is reducing the variance of P_0, without generating any more samples from either population. How would the estimate of P_0 change if it were based on the negatives of the (pseudo-) random samples? Since T and t are symmetric, $T(-\underline{X})$ (resp. $t(-\underline{x})$) estimates $-\mu_1$ (resp. $-\mu_2$). Therefore, evaluating the conditional probability in (1) based on $-X_1,\dots,-X_{n_1}$ and $-x_1,\dots,-x_{n_2}$ yields

$$\underset{\underline{C},\underline{c},\underline{X},\underline{x}}{E} \Phi \left[\frac{k(S^2_{T(\underline{C})} + s^2_{t(\underline{c})})^{1/2} + T(\underline{C}) - t(\underline{c})}{(1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} y_i^2)^{1/2}} \right] \stackrel{\Delta}{=} \frac{1}{N} \cdot \sum_{j=1}^{N} \Phi_{2j} \quad (2)$$

where

$$\Phi_{2j} = \Phi \left[\frac{k(S_j^2 + s_j^2)^{1/2} + (T_j - A_j) - (t_j - a_j)}{(1/\sum_{i=1}^{n_1} Y_{ij}^2 + 1/\sum_{i=1}^{n_2} y_{ij}^2)^{1/2}} \right]$$

Now $P_0\{W \le k\}$ can be estimated by

$$P_0 = \frac{1}{N} \cdot \sum_{j=1}^{n} Y_j ,$$

where

$$Y_j = \frac{1}{2}(\phi_{1j} + \phi_{2j}) .$$

Again,

$$Var(\hat{P}_0) \overset{\Delta}{=} \frac{1}{N}\left[\frac{1}{N}\sum_{j=1}^{N} Y_j^2 - (\hat{P}_0)^2 \right]$$

but now the estimate of $\hat{P}_0$ has a smaller variance, since each Y_j involves two quantities which are negatively correlated. Gross ([1]) derived this improvement by recognizing that $P_0\{W \leq k\} = P_0\{W \geq -k\}$ by the symmetry of W, and (2) is a direct estimate of this latter probability.

Further reductions in the variance of $\hat{P}_0$ are possible if the symmetry of the distribution of t, say $G(z - \mu_2)$, is assumed. Just as there is no loss of generality in taking the difference between μ_1 and μ_2 to be 0 in the null hypothesis (the distribution of W is a function of the difference $(\mu_1 - \mu_2)$ only), there will be no loss of generality in taking μ_2 to be 0, and the free parameter is μ_1. Furthermore, the symmetry assumption on the distribution of t says that

$$G(\mu_2 + z) = 1 - G(\mu_2 - z)$$

i.e.,

$$G(z) = 1 - G(-z)$$

$$\Longrightarrow dG(z) = -dG(-z) .$$

Let $F(z - \mu_1)$ denote the distribution of T, and let

$S_{samp} = \sqrt{S^2_{T(\underline{X})} + s^2_{t(\underline{x})}}$ and $S_{res} = \sqrt{S^2_{T(\underline{C})} + s^2_{t(\underline{c})}}$. Then we have that

$$\begin{aligned}
P\{W \leq k\} &= P\{(T-t)/S_{samp} \leq k\} \\
&= P\{T \leq kS_{samp} + t\} \\
&= \int_{-\infty}^{\infty} P_0\{T \leq kS_{samp} + z\} dG(z) \\
&= \int_{-\infty}^{\infty} F(kS_{samp} + z - \mu_1)\, dG(z) \\
&= -\int_{-\infty}^{\infty} F(kS_{samp} + z - \mu_1)\, dG(-z) \\
&= \int_{-\infty}^{\infty} F(kS_{samp} - v - \mu_1)\, dG(v) \\
&= \int_{-\infty}^{\infty} P\{T \leq kS_{samp} - v\}\, dG(v) \\
&= P\{T + t \leq kS_{samp}\} \\
&= P\{T(\underline{X}) + t(\underline{x}) \leq kS_{samp}\} \\
&= P\{T(\underline{C}) + A + t(\underline{c}) + a \leq kS_{res}\} \\
&= E_{\underline{C},\underline{c},\underline{X},\underline{x}}\, P\{A + a \leq kS_{res} - T(\underline{C}) - t(\underline{c}) \mid \underline{C},\underline{c},\underline{X},\underline{x}\} .
\end{aligned}$$

Under the null hypothesis,

$$A + a \sim N(\mu_1 + 0, \frac{1}{\Sigma Y_i^2} + \frac{1}{\Sigma y_i^2}).$$

Hence,

$$P_0\{W \leq k\} = E_{\underline{C},\underline{c},\underline{Y},\underline{y}}\, \Phi\left[\frac{kS_{res} - T(\underline{C}) + t(\underline{c})}{\left[1/\sum_{i=1}^{n_1} Y_i^2 + 1/\sum_{i=1}^{n_2} y_i^2\right]}\right]$$

$$\underset{=}{\Delta} \frac{1}{N} \cdot \sum_{j=1}^{N} \Phi_{3j}$$

where

$$= \Phi_{3j} = \Phi\left[\frac{k\sqrt{S_j^2 + s_j^2} - (T_j - A_j) - (t_j - a_j)}{\sqrt{1/\sum_{i=1}^{n_1} Y_{ij}^2 + 1/\sum_{i=1}^{n_2} y_{ij}^2}}\right]$$

Again, recognizing the symmetry of the conditional distribution of $A + a$, we arrive at

$$P_0\{W \leq k\} \underset{=}{\Delta} \frac{1}{N} \cdot \sum_{j=1}^{n} \Phi_{4j} \ ,$$

where

$$\Phi_{4j} = \Phi\left[\frac{k\sqrt{S_j^2 + s_j^2} + (T_j - A_j) + (t_j - a_j)}{\left[1/\sum_{i=1}^{n_1} Y_{ij}^2 + 1/\sum_{i=1}^{n_2} y_{ij}^2\right]^{1/2}}\right].$$

In taking our final estimate of P_0 as

$$\frac{1}{N} \cdot \sum_{j=1}^{N} Y_j \ , \ Y_j = \frac{1}{4}(\Phi_{1j} + \Phi_{2j} + \Phi_{3j} + \Phi_{4j}) \qquad (3)$$

it remains to check that all six pairs between the Φ_{kj} are negatively correlated. It is clear that the two pairs Φ_{1j}, Φ_{2j} and Φ_{3j}, Φ_{4j} are, since they are based on antithetic sampling. That Φ_{1j}, Φ_{3j} and Φ_{2j}, Φ_{4j} are negatively correlated is shown by the fact that if $T(\underline{C})$ is held fixed, Φ_{1j} (resp. Φ_{4j}) is increasing in $t(\underline{c})$, while Φ_{3j} (resp. Φ_{2j}) decreases in $t(\underline{c})$. Likewise, holding $t(\underline{c})$ fixed, Φ_{1j}, Φ_{4j} and Φ_{2j}, Φ_{3j} are negatively correlated by the same reasoning on $T(\underline{C})$. Hence, the overall variance of Y_j has been reduced, and we can use (3) as our final estimate.

One further improvement may be worth investigating. As in equation (1),

$$P_0\{W \leq k\} = E_{\underline{C},\underline{c},\underline{Y},\underline{y}}\, P\{A - a \leq kS_{res} - T(C) + t(c) \mid \underline{C},\underline{c},\underline{Y},\underline{y}\}$$

$$= E_{\underline{C},\underline{c},\underline{Y},\underline{y}}\, P\{A \leq kS_{res} - T(C) + t(c) + a \mid \underline{C},\underline{c},\underline{Y},\underline{y}\}$$

Given $\underline{C},\underline{c},\underline{Y},\underline{y}$, the distribution of A <u>alone</u> is $N(\mu_1, 1/\Sigma Y_i^2) = N(0, 1/\Sigma Y_i^2)$, and the conditional distribution of a is $N(\mu_2, 1/\Sigma y_i^2) = N(0, 1/\Sigma y_i^2)$ (under the null hypothesis). The inner conditional probability can then be evaluated as

$$P(A \le kS_{res} - T(C) + t(c) + a \mid \underline{C}, \underline{c}, \underline{X}, \underline{x})$$

$$= \int_{-\infty}^{\infty} P(A \le kS_{res} - T(\underline{C}) + t(\underline{c}) + z \mid \underline{C}, \underline{c}, \underline{X}, \underline{x}) \; d\Phi(z \sqrt{\Sigma y_i^2})$$

$$= \int_{-\infty}^{\infty} \Phi[(kS_{res} - T(C) + t(c) + z)\sqrt{\Sigma Y_i^2}] \; d\Phi(z \sqrt{\Sigma y_i^2})$$

$$= (\Sigma y_i^2)^{-1/2} \int_{-\infty}^{\infty} \Phi\left[\frac{u - (\Sigma y_i^2)^{1/2}(-kS_{res} + T(C) - t(c))}{(\Sigma y_i^2 / \Sigma Y_i^2)^{1/2}}\right] d\Phi(u)$$

$$= (\Sigma y_i^2)^{-1/2} \int_{-\infty}^{\infty} \Phi\left[\frac{u - \xi}{\eta}\right] d\Phi(u)$$

where

$$\xi = (\Sigma y_i^2)^{1/2}(-kS_{res} + T(C) - t(c))$$

$$\eta^2 = \sum_{i=1}^{n_2} y_i^2 \Big/ \sum_{i=1}^{n_1} Y_i^2 \quad .$$

An evaluation of this integral is numerically possible (see Note 1). The result will be an increasing function of

$$\xi = (\Sigma y_i^2)^{1/2}(-kS_{res} + T(C) - t(c)).$$

Thus,

$$\Phi_{5j} = (\Sigma y_{ij}^2)^{-1/2} \; .$$

$$\int_{-\infty}^{\infty} \Phi\left[\frac{u - (\Sigma y_{ij}^2)^{1/2}[-k(S_j^2 + s_j^2)^{1/2} + (T_j - A_j) - (t_j - a_j)]}{(\Sigma y_{ij}^2 / \Sigma Y_{ij}^2)^{1/2}}\right] d\Phi(u)$$

will be negatively correlated with each of the other Φ_{kj}, $k = 1, \ldots, 4$. However, the additional computation required to

estimate $P_0(W \le k)$ by

$$P_0 = N^{-1} \cdot \sum_{j=1}^{N} Y_j \; , \qquad Y_j = (\Phi_{1j} + \cdots + \Phi_{5j})/5$$

may not be worth the possibly slight reduction in the variance of P_0.

B. Location and Scale Swindle.

The location and scale swindle follows the same principle as the location swindle, but now the configurations are taken as

$$T(\underline{C}) = T\left(\frac{\underline{X} - A}{B}\right) = \frac{T(\underline{X}) - A}{B}$$

and

$$t(\underline{c}) = t\left(\frac{\underline{x} - a}{B}\right) = \frac{t(\underline{x}) - a}{B} \; ,$$

where

$$(n_1 + n_2 - 2)B^2 = \sum_{i=1}^{n_1} (Z_i - AY_i)^2 + \sum_{i=1}^{n_2} (z_i - ay_i)^2 \quad .$$

As illustrated in [4], the first term is distributed as $\chi^2_{(n_1-1)}$ and is independent of A; since the second term is based on an entirely different sample, B itself is independent of A. Likewise, the second term is distributed as $\chi^2_{(n_2-1)}$; by the same reasoning, B is also independent of a. By the reproductive property of the chi-squared distribution

$$(n_1 + n_2 - 2)B^2 \sim \chi^2_{(n_1+n_2-2)} \cdot$$

Now,

$$S_T^2 = \mathrm{Var}(T(\underline{X})) = \mathrm{Var}\left[\left(\frac{T(\underline{X})-S}{B} \right) \cdot B + A \right]$$

$$= \mathrm{Var}(T(\underline{C})B + A) = B^2 \mathrm{Var}(T(\underline{C})) = B^2 S_C^2$$

and similarly

$$s_t^2 = B^2 s_c^2 .$$

Hence,

$$S_{samp} = (S_T^2 + s_t^2)^{1/2} = BS_{res} .$$

The location and scale swindle then proceeds as:

$$P_0\{W \le k\} = P_0\{T(\underline{X}) - t(\underline{x}) \le k \cdot S_{samp}\}$$

$$= P_0\{ [(\tfrac{T(\underline{X})-A}{B})B + A] - [(\tfrac{t(\underline{x})-a}{b})b + a] \le kS_{samp} \}$$

$$= P_0\{ T(\underline{C})B - t(\underline{c})B \le kBS_{res} - A + a \}$$

$$= E_{\underline{C},\underline{c},\underline{X},\underline{x}} P_0\{ \frac{kS_{res}B - A + a}{B} \ge T(\underline{C}) - t(\underline{c}) \mid \underline{C},\underline{c},\underline{X},\underline{x} \}$$

$$= E_{\underline{C},\underline{c},\underline{X},\underline{x}} P_0\{ kS_{res} - \frac{A-a}{B} \ge T(\underline{C}) - t(\underline{c}) \mid \underline{C},\underline{c},\underline{X},\underline{x} \}$$

$$= E_{\underline{C},\underline{c},\underline{Y},\underline{y}} P_0\left[\frac{A-a}{B\left[\frac{1}{\Sigma Y_i^2} + \frac{1}{\Sigma y_i^2}\right]^{1/2}} \le \frac{kS_{res} - (T(\underline{C}) - t(\underline{c}))}{\left[\frac{1}{\Sigma Y_i^2} + \frac{1}{\Sigma y_i^2}\right]^{1/2}} \mid \underline{C},\underline{c},\underline{Y},\underline{y} \right]$$

The expression

$$\frac{kS_{res} - (T(\underline{C}) - t(\underline{c}))}{(1/\Sigma Y_i^2 + 1/\Sigma y_i^2)^{1/2}}$$

depends only on the configuration and on $\underline{Y},\underline{y}$. Also, since the conditional distribution of the numerator is

$$\frac{A-a}{(1/\Sigma Y_i^2 + 1/\Sigma y_i^2)^{1/2}} \sim N(0,1)$$

and that of the denominator is

$$B \sim \sqrt{\chi^2_{(n_1+n_2-2)} / (n_1+n_2-2)} ,$$

the conditional probability is Student's t on $(n_1 + n_2 - 2)$ degrees of freedom. Thus,

$$P_0\{W \le k\} \triangleq \frac{1}{N} \sum_{j=1}^{N} F_\nu \left[\frac{k(S_{C_j}^2 + s_{c_j}^2)^{1/2} - (\frac{T_j - A_j}{B_j}) + (\frac{t_j - a_j}{B_j})}{(1/\Sigma Y_i^2 + 1/\Sigma y_i^2)^{1/2}} \right]$$

where $F_\nu(\cdot)$ is the cumulative distribution function of Student's-t on $\nu = n_1 + n_2 - 2$ degrees of freedom. As in the location swindle, the same improvements in the estimate of P_0 by using all possible combinations of sign on $T(\underline{C})$ and $t(\underline{c})$ are applicable here as well.

C. Power Estimates.

Using either swindle, the power of W against various points in the alternative, H_1: $\mu_1 - \mu_2 = \delta$, can be approximated. For the location swindle,

$$P_\delta\{W \le k\} = P_\delta\{\, T(\underline{X}) - t(\underline{x}) \ge kS_{samp} \,\}$$

$$= P_\delta\{\, (T(\underline{C})+A) - (t(\underline{c})-a) \ge kS_{res} \,\}$$

$$= E_{\underline{C},\underline{c},\underline{Y},\underline{y}} P_\delta\{\, A-a-\delta \ge kS_{res} - T(\underline{C}) + t(\underline{c}) - \delta \mid \underline{C},\underline{c},\underline{Y},\underline{y} \}$$

$$\triangleq \frac{1}{N} \sum_{j=1}^{N} \Phi \left[\frac{(T_j - A_j) - (t_j - a_j) + \delta - kS_j}{\left(1/\sum_{i=1}^{n_1} Y_{ij}^2 + 1/\sum_{i=1}^{n_2} y_{ij}^2\right)^{1/2}} \right]$$

where again all four possible combinations of sign on $(T_j - A_j)$ and $(t_j - a_j)$ may be utilized to achieve a more efficient estimate.

In the case of the location and scale swindle, however, the conditional probability is <u>noncentral</u> Student's t:

$$P_\delta\{W \le k\} = P_\delta\{\, (T(\underline{C})B+A) - (t(\underline{c})B+a) \ge kBS_{res} \,\}$$

$$= P_\delta\{\, \frac{A-a}{B} \ge kS_{res} - T(\underline{C}) + t(\underline{c}) \,\}$$

$$= E_{\underline{C},\underline{c},\underline{Y},\underline{y}} P_\delta \left[\frac{A-a}{B(1/\sum Y_i^2 + 1/\sum y_i^2)^{1/2}} \ge \frac{kS_{res} - T(\underline{C}) + t(\underline{c})}{(1/\sum Y_i^2 + 1/\sum y_i^2)^{1/2}} \mid \underline{C},\underline{c},\underline{Y},\underline{y} \right]$$

$$\triangleq N^{-1} \sum_{j=1}^{N} \left[1 - F_{\delta,\nu}\left[\frac{kS_{res} - T(\underline{C}) + t(\underline{c})}{(1/\sum Y_i^2 + 1/\sum y_i^2)^{1/2}} \right] \right]$$

where $F_{\delta,\nu}$ is a tabulated or approximated noncentral-t with noncentrality parameter δ and degrees of freedom $\nu = n_1 + n_2 - 2$.

An alternative approach to achieving power estimates that does not require a noncentral Student's t would be to swindle on scale separately as follows:

Suppose $X \sim N(\delta, 1)$, $B \sim \sqrt{\chi_n^2/n}$. Then

$$P\{\, \frac{X}{B} \ge r \,\} = 1 - F_{\delta,n}(r) ,$$

$$= \int_0^\infty P\,\{X \ge Br \mid B=b\}\, dF^B(b) .$$

Since $g(B) = nB^2 \sim \chi_n^2$, $\partial g/\partial B = 2nB$,

$$f^B(b) = (\Gamma(n/2) \cdot 2^{n/2})^{-1} \cdot (nb^2)^{n/2 - 1} \cdot e^{-nb^2/2} \cdot I_{(0,\infty)}(b)$$

$$= n^{n/2} (\Gamma(n/2) \cdot 2^{n/2})^{-1} \cdot b^{n-1} \cdot e^{-nb^2/2} \cdot I_{(0,\infty)}(b) .$$

Thus,

$$P\left(\frac{X}{B} \geq r\right) = \int_0^{\infty} P\left(X - \delta \geq br - \delta\right) dF^B(b)$$

$$= \int_0^{\infty} \Phi(\delta - br)\cdot n^{n/2}(\Gamma(n/2)\cdot 2^{n/2})^{-1}\cdot b^{n-1}\cdot e^{-nb^2/2}\, db \ .$$

Evaluation of the integral is numerically possible. (The details will not be shown here, but the argument follows the same lines as that in Note 1.) Since r represented

$$\frac{kS_{res} - T(\underline{C}) - t(\underline{c})}{(1/\Sigma Y_i^2 + 1/\Sigma y_i^2)^{1/2}} \ ,$$

additional efficiency can again be obtained by using all possibility of sign on $T(\underline{C})$ and $t(\underline{c})$.

__Note 1.__ __Evaluation of__ $\rho\cdot \int_{-\infty}^{\infty} \Phi\left(\frac{u-\xi}{\eta}\right) d\Phi(u)$, where

$$\rho = \left(\sum_{i=1}^{n_2} y_i^2\right)^{-1/2}$$

$$\xi = \rho^{-1}\cdot\left(-kS_{res} + T(\underline{C}) - t(\underline{c})\right)$$

$$\eta^2 = \sum_{i=1}^{n_2} y_i^2 \Big/ \sum_{i=1}^{n_2} Y_i^2 \qquad .$$

First, let us note that:

(i) as $\xi/\eta \to -\infty$, $\Phi\left(\frac{u-\xi}{\eta}\right)\cdot\phi(u) \to 0$

$\Rightarrow \int\Phi\left(\frac{u-\xi}{\eta}\right)\cdot d\Phi(u) \to 0$;

(ii) for $\xi = 0$, $\eta = 1$, $\int\Phi(u)d\Phi(u) = 1/2$;

(iii) as $\xi/\eta \to +\infty$, $\Phi\left(\frac{u-\xi}{\eta}\right)\phi(u) \doteq \Phi\left(\frac{u-\xi}{\eta}\right)\cdot 1$

$\to \int\Phi\left(\frac{u-\xi}{\eta}\right)d\Phi(u) \to 1$.

The "sledgehammer" approach expands $\Phi\left(\frac{u-\xi}{\eta}\right)$ in a Taylor series about 0 and uses the moment properties of the Gaussian distribution:

$$\int_{-\infty}^{\infty} \Phi\left(\frac{u-\xi}{\eta}\right) d\Phi(u) = \int_{-\infty}^{\infty}\left[\sum_{k=0}^{\infty}\left(\frac{u-\xi}{\eta}\right)^k \frac{1}{k!}\phi^{(k)}(0)\right] d\Phi(u)$$

$$= \sum_{k=0}^{\infty} \frac{\phi^{(k)}(0)}{\eta^k k!} \int_{-\infty}^{\infty} (u - \xi)^k \, d\Phi(u)$$

(interchange OK: everything absolutely integrable)

$$= \sum_{k=0}^{\infty} \frac{\phi^{(k)}(0)}{\eta^k k!} \sum_{j=0}^{k} \binom{k}{j}(\xi)^{k-j} \int_{-\infty}^{\infty} u^j d\Phi(j)$$

$$= \frac{1}{2} + \sum_{p=0}^{\infty} \frac{\phi^{(p)}(0)}{\eta^{p+1}(p+1)!} \sum_{j=0}^{p+1} \binom{p+1}{j}(-\xi)^{p+1-j} E(N^j)$$

Now,

$$\phi^{(p)}(0) = \begin{cases} 0 & p \text{ odd} \\ \dfrac{(-1/2)^{p/2}(p)!}{(2\pi)^{1/2}(p/2)!} & p \text{ even} \end{cases}$$

and

$$E(N^j) = \begin{cases} 0 & j \text{ odd} \\ \dfrac{j!}{2^{j/2}(j/2)!} & j \text{ even} \end{cases} \quad .$$

Substituting in these formulas, and changing the index of summation,

$$\int_{-\infty}^{\infty} \Phi\left(\frac{u-\xi}{\eta}\right) d\Phi(u) = \frac{1}{2} + \sum_{m=0}^{\infty} \frac{(-1)^m (2m)!}{2^m m!} \sum_{p=0}^{m} \frac{(-\xi/\eta)^{2(m-p)+1}}{[2(m-p)+1]!2^p p!}$$

which, unfortunately, looks difficult in providing a numerical answer.

Another more feasible approach is in terms of standardized Hermite polynomials. Following [2], the integral can be written as

$$\int \Phi(u) d\Phi\left(\frac{u-\xi}{\eta}\right) = \int \Phi(u) \phi_{\xi,\eta^2}(u) du$$

$$\xi = \left(\sum_{i=1}^{n_1}\right)^{1/2} (k - T(\underline{C}) + t(\underline{c})) \ ; \quad \eta^2 = \sum_{i=1}^{n_1} Y_i^2 \Big/ \sum_{i=1}^{n_2} y_i^2$$

Now,

$$\int \Phi(u) \phi_{\xi,\eta^2}(u) du = \int \Phi_{0,1}(u) \cdot \frac{\phi_{\xi,\eta^2}(u)}{\phi_{0,1}(u)} \cdot \phi_{0,1}(u)$$

$$= \left\langle \Phi_{0,1} , \frac{\phi_{\xi,\eta^2}}{\phi_{0,1}} \right\rangle$$

where $\langle \cdot , \cdot \rangle$ is the inner product in the L^2 space of functions spanned by $\phi_{0,1}(Y)dY$ as basis elements. Using the standardized Hermite polynomials

$$h_n(x) = (n!)^{-1/2} He_n(x) ,$$

this inner product may be written as

$$\sum_{p=0}^{\infty} \langle \Phi_{0,1} , h_p \rangle \cdot \left\langle \frac{\phi_{\xi,\eta^2}}{\phi_{0,1}} , h_p \right\rangle .$$

Using Equation (3.7) from [2], $\langle \Phi_{0,1} , h_p \rangle$ may be expressed as

$$\langle \Phi_{0,1} , h_p \rangle = \begin{cases} \dfrac{2(-1)^{(p-1)/2}(p!)^{1/2}}{\sqrt{\pi}((p-1)/2)!3^{(p+1)/4}p} & p \text{ odd} \\ 0 & p \text{ even} \end{cases}$$

and, using (5.4) and (5.5) from [2],

$$\left\langle \frac{\phi_{\xi,\eta^2}}{\phi_{0,1}} , h_p \right\rangle = \begin{cases} (1-\eta^2)^{p/2} h_p(\xi(1-\eta^2)^{-1/2}) & \eta^2 < 1 \\ \xi^1 / 1! & \eta = 1 \end{cases}$$

Thus,

$$\int \Phi_{0,1}(x) d\Phi\left(\frac{x-\xi}{\eta^2}\right) = \frac{2}{\sqrt{\pi}} \sum_{p=0}^{\infty} \frac{(-1)^{(p-1)/2}(p!)^{1/2}}{((p-1)/2)!3^{(p+1)/4}p} \cdot (1-\eta^2)^{p/2}$$

$$\cdot h_p(\xi(1-\eta^2)^{-1/2} \qquad , p \text{ odd}, \ \eta^2 < 1$$

$$= \frac{2}{\sqrt{\pi}} \sum_{r=0}^{\infty} \frac{(-1)^r((2r+1)!)^{1/2}}{r!3^{(r+1)/2}(2r+1)} \cdot (1-\eta^2)^{r+1/2}$$

$$\cdot h_{2r+1}(\xi(1-\eta^2)^{-1/2}) \quad , \ \eta^2 < 1$$

$$= \frac{2}{\sqrt{\pi}} \sum_{r=0}^{\infty} \frac{(-1)^r((2r+1)!)^{1/2}}{r!3^{(r+1)/2}(2r+1)} \cdot \xi^{2r+1} \cdot ((2r+1)!)^{-1/2}$$

$$= \frac{2}{\sqrt{\pi}} \sum_{r=0}^{\infty} \frac{(-1)^r \xi^{2r+1}}{r!3^{(r+1)/2}(2r+1)} \qquad , \ \eta^2 = 1 .$$

The restriction $\eta^2 \leq 1$ can be assured in the Gaussian case, either when the sample sizes are equal ($\Sigma Y_i^2 = \Sigma y_i^2 = n$) or with unequal sample sizes, if population 2 is the one having the larger sample size ($n_1 < n_2$). When the distributions are different, no guarantee can be made that $\eta^2 \leq 1$; the samples would have to be compared in each case so that ($\sum_{i=1}^{n_1} Y_{ij}^2 = \sum_{i=1}^{n_2} y_{ij}^2$) , for all $j=1,\cdots,N$ = # of samples. For anything other than the Gaussian situation, this is likely to be more trouble than it is worth. For the Gaussian situation, tabulated values of $n_{2r+1}(\xi(1 - n_1/n_2)^{-1/2})$ may assist in approximating the integral, from which a suitable cutoff limit in the infinite summation may be derived to obtain a sufficiently accurate finite approximation. (To date, no such results were attempted; hence, the improvement in the variance of the estimate of P_0 has not been determined.)

REFERENCES

[1] Gross, A. M. (1973). A Monte Carlo Swindle for Estimators of location. *Appl. Statist.*, *22*, No. 3, 347-353.

[2] Kates, Louis K. (1977). Hermite Polynomials and the Gaussian Distribution. Unpublished manuscript, Princeton University.

[3] Relles, D. (1970). Variance Reduction Techniques for Monte Carlo Sampling from Student Distributions. *Technometrics*, *12*, 499-515.

[4] Simon, Gary (1975). Swindles to Improve Computer Simulation, with Application of the Problems of Appraising Estimates of Location and Dispersion in Univariate Samples. Technical Report No. 91, Series 2, Princeton University.